



## OPEN ACCESS

EDITED BY  
Quinn McFrederick,  
University of California, Riverside,  
United States

REVIEWED BY  
Kirk Anderson,  
Pacific West Area, Agricultural  
Research Service (USDA), United States  
Kasie Raymann,  
University of North Carolina  
at Greensboro, United States

\*CORRESPONDENCE  
Shiqi Luo  
shiqi\_luo@cau.edu.cn  
Xin Zhou  
xinzhou@cau.edu.cn

†These authors have contributed  
equally to this work

SPECIALTY SECTION  
This article was submitted to  
Microbial Symbioses,  
a section of the journal  
Frontiers in Microbiology

RECEIVED 02 May 2022  
ACCEPTED 29 July 2022  
PUBLISHED 02 September 2022

CITATION  
Su Q, Tang M, Hu J, Tang J, Zhang X,  
Li X, Niu Q, Zhou X, Luo S and Zhou X  
(2022) Significant compositional  
and functional variation reveals  
the patterns of gut microbiota  
evolution among the widespread Asian  
honeybee populations.  
*Front. Microbiol.* 13:934459.  
doi: 10.3389/fmicb.2022.934459

COPYRIGHT  
© 2022 Su, Tang, Hu, Tang, Zhang, Li,  
Niu, Zhou, Luo and Zhou. This is an  
open-access article distributed under  
the terms of the [Creative Commons  
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other  
forums is permitted, provided the  
original author(s) and the copyright  
owner(s) are credited and that the  
original publication in this journal is  
cited, in accordance with accepted  
academic practice. No use, distribution  
or reproduction is permitted which  
does not comply with these terms.

# Significant compositional and functional variation reveals the patterns of gut microbiota evolution among the widespread Asian honeybee populations

Qinzhi Su<sup>1,2†</sup>, Min Tang<sup>2†</sup>, Jiahui Hu<sup>2</sup>, Junbo Tang<sup>1,2</sup>,  
Xue Zhang<sup>2</sup>, Xingan Li<sup>3</sup>, Qingsheng Niu<sup>3</sup>, Xuguo Zhou<sup>4</sup>,  
Shiqi Luo<sup>2\*</sup> and Xin Zhou<sup>2\*</sup>

<sup>1</sup>College of Food Science and Nutritional Engineering, China Agricultural University, Beijing, China, <sup>2</sup>Department of Entomology, College of Plant Protection, China Agricultural University, Beijing, China, <sup>3</sup>Key Laboratory for Bee Genetics and Breeding, Jilin Provincial Institute of Apicultural Sciences, Jilin, China, <sup>4</sup>Department of Entomology, University of Kentucky, Lexington, KY, United States

The gut microbiome is a crucial element that facilitates a host's adaptation to a changing environment. Compared to the western honeybee *Apis mellifera*, the Asian honeybee, *Apis cerana* populations across its natural range remain mostly semi-feral and are less affected by bee management, which provides a good system to investigate how gut microbiota evolve under environmental heterogeneity on large geographic scales. We compared and analyzed the gut microbiomes of 99 Asian honeybees, from genetically diverged populations covering 13 provinces across China. Bacterial composition varied significantly across populations at phylotype, sequence-discrete population (SDP), and strain levels, but with extensive overlaps, indicating that the diversity of microbial community among *A. cerana* populations is driven by nestedness. Pollen diets were significantly correlated with both the composition and function of the gut microbiome. Core bacteria, *Gilliamella* and *Lactobacillus* Firm-5, showed antagonistic turnovers and contributed to the enrichment in carbohydrate transport and metabolism. By feeding and inoculation bioassays, we confirmed that the variations in pollen polysaccharide composition contributed to the trade-off of these core bacteria. Progressive change, i.e., nestedness, is the foundation of gut microbiome evolution among the Asian honeybee. Such a transition during the co-diversification of gut microbiomes is affected by environmental factors, diets in general, and pollen polysaccharides in particular.

## KEYWORDS

gut microbiota, Asian honeybee, population variation, pollen, nectar

## Introduction

The gut microbiome often serves as a critical component in the host's adaptation to a changing environment (Suzuki and Ley, 2020). Gut microbiota can benefit host animals in nutrition provision, pathogen resistance, and modulations of development and behavior (Cryan and Dinan, 2012; Engel and Moran, 2013; Kamada et al., 2013; Sommer and Bäckhed, 2013). On the other hand, gut microbiota may be shaped by the host's adjustments to changing environments, such as range expansion accompanied by diet shifts (Baldo et al., 2015; Michel et al., 2018). In particular, for widespread species found in a large geographic range, environmental heterogeneity is expected to influence their gut microbiota (Yatsunenko et al., 2012; Henderson et al., 2015). This is because the geographic location of animal populations is linked with varied host genetics, local vegetation, and environmental microbe sources.

Studies based on *Apis mellifera* have established the framework for honeybee gut microbiota, revealing their essential role in the biology of the honeybee, such as facilitating pollen digestion (Engel et al., 2012; Zheng et al., 2019), host development (Zheng et al., 2017), and pathogen resistance (Kwong et al., 2017a; Wu et al., 2020). The species of honeybees each maintain a relatively simple but stable gut microbiota, comprising 5–9 core bacteria (>95% of total abundance) from phyla Proteobacteria, Firmicutes, and Actinobacteria (Kwong and Moran, 2016; Kwong et al., 2017b). Ancestry reconstruction of these core microbes suggested that they have probably become part of the symbiont system in the common ancestry of all extent corbiculate bees (Kwong et al., 2017b). Interestingly, although the honeybees share much of the core microbes at the phylotype level, each host species possesses a species-specific microbial community (Kwong et al., 2017b), with most core microbes showing distinct strain diversities among hosts, e.g., between *Apis mellifera* and *Apis cerana* (Ellegaard et al., 2020). However, little is known about how these gut symbionts have evolved within their hosts.

Among the different honeybee species, both the western (*A. mellifera*) and eastern honeybees (*A. cerana*) are widely distributed across tropical and temperate climates, each with endemic populations adapted to local habitats (Wallberg et al., 2014; Ji et al., 2020). Compared to *A. mellifera*, *A. cerana* populations across its natural range (much of eastern, southern, and southeastern Asia) (Radloff et al., 2010) remain mostly semi-feral and are less affected by bee management, which provides a good system to investigate how gut microbiota evolve under environmental heterogeneity on large geographic scales. However, investigations based on 16S rRNA did not provide sufficient resolution to differentiate tropical *A. cerana* populations from those of the temperate zones (Kwong et al., 2017b). There is still a knowledge gap on the biogeographical variation of gut microbes among *A. cerana* populations.

Our recent work on the evolution of mainland *A. cerana* revealed that multiple peripheral subspecies had radiated from a common central ancestral population and adapted independently to diverse habitats (Ji et al., 2020). During the most recent radiation period (~100 ka), selective pressures imposed by diverse habitats, especially those of the changing floras, led to the convergent adaptation of the honeybee, where genes associated with sucrose sensitivity and foraging labor division had been repeatedly selected (Ji et al., 2020). We hypothesized that the gut microbiota of *A. cerana* had also evolved along with host range expansion, subspecies differentiation, and habitat adaptation. In the present study, we aimed to understand the landscape of gut microbial diversity and function across geographic populations of mainland *A. cerana* with metagenome sequencing. We also examined the effects of host genetics and diet variation on the honeybee gut symbionts. In addition, we explored the adaptive mechanisms of the microbes in response to selective pressures.

## Materials and methods

### Sample collection

A total of 99 worker bees of *A. cerana* were obtained from inside the hives at 15 sites in 13 provinces of China (Hainan, Yunnan, Taiwan, Fujian, Jiangxi, Hunan, Tibet, Sichuan, Shaanxi, Gansu, Qinghai, Hebei, and Jilin), between April 2017 and January 2019. For each population,  $\geq 5$  gut samples were sequenced from at least two hives to represent the diversity of each population (Supplementary Table 1). We collected nurse bees based on their morphology (Seeley, 1982). The nurse bees are generally characterized by relatively lightened color and apparently intact hairs and wings. Bees that were newly emerged (with shiny hair and slow-moving capability) or aged (with visible wing wear and hair loss) were excluded from sampling. Our sampling covered the main natural distribution range of *A. cerana* in China, from 19.2°–43.5°N, 95.7°–128.7°E, representing drastically different altitudes (12–3,325 m, Supplementary Table 1). The guts (including the midgut and hindgut) were dissected from the abdomen and stored in 100% ethanol or directly frozen at –80°C. To preserve live gut bacteria for strain isolation, a subset of guts was suspended in 100  $\mu$ l of 25% glycerol (v/v, dissolved in PBS buffer), homogenized, and then frozen at –80°C.

### Isolation, cultivation, and identification of gut microbe strains

The gut homogenates were plated on different cultivation media, respectively, for various honeybee gut bacteria following Engel et al. (2013), including heart infusion agar (HIA) with 5%

(v/v) de-fibrinated sheep blood, Columbia agar with 5% (v/v) de-fibrinated sheep blood, De Man, Rogosa and Sharpe (MRS) agar, and trypticase–phytone–yeast (TPY) agar supplemented with 1% mupirocin. The plates were incubated at 35°C in a 5% CO<sub>2</sub> or anaerobic atmosphere.

When bacterial colonies became visible on the plates, they were identified by sequences of their 16S rRNA gene. The isolates were picked and dissolved with H<sub>2</sub>O, then boiled at 100°C for 1 min, which was used directly as a DNA template in PCR. PCR amplicons were generated using the universal 16S primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492R (5'-GGTACCTTGTTACGACTT-3') with 25 cycles of amplification (94°C for 30 s, 60°C for 40 s, and 72°C for 60 s) after an initial incubation for 1 min at 95°C. Amplicons were sequenced using Sanger sequencing and identified using blastn against annotated sequences in GenBank.

## DNA extraction for genome and metagenome sequencing

The gut DNA was extracted following Kwong et al. (2017b). Briefly, the crushed gut was suspended in a capped tube with 728 µl of CTAB buffer, 20 µl of proteinase K, 500 µl of 0.1-mm Zirconia beads (BioSpec), 2 µl of 2-Mercaptoethanol, and 2 µl of RNase A cocktail. The mixtures were bead-beaten for 2 min for 3 times. After digested overnight at 50°C, the mixtures were added with 750 µl of phenol/chloroform/isoamyl alcohol (25:24:1, pH 8.0) and centrifuged to obtain the aqueous layer. After being precipitated at -20°C, spun at 4°C, and washed with -20°C ethanol, the DNA pellets were dried at 50°C, and then re-suspended in 50 µl of nuclease-free H<sub>2</sub>O. Final DNA samples were stored at -20°C.

Genomic DNA of honeybee gut bacterial isolates was also extracted using the phenol-chloroform protocol. The bacterial cells were re-suspended in 500 µl of lysis buffer [50 mM Tris-HCl (pH 8.0), 200 mM NaCl, 20 mM EDTA, nuclease-free H<sub>2</sub>O, 2% SDS, proteinase K (20 mg/ml)], then added with 500 µl of CTAB extraction buffer [50 mM Tris-HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, 2% CTAB, 1% PVP 40000, nuclease-free H<sub>2</sub>O; pre-heated at 56°C]. The mixtures were incubated for 30 min at 65°C before the addition of 500 µl of phenol/chloroform/isoamyl alcohol (25:24:1, pH 8.0). Then, the mixture was centrifuged at 14,000 g at room temperature (RT) for 5 min. The aqueous layer was transferred to a new tube, added with 5 µl of RNase (100 mg/ml), incubated at RT for 20 min, and added with 600 µl of chloroform: isoamyl alcohol (24:1). After spinning at 14,000 g at RT for 5 min, the aqueous layer was transferred to a new tube and added with 5 µl of ammonium acetate (final concentration 0.75 M), 1 µl glycogen solution (20 mg/ml), and 1 ml of cold 100% ethanol. DNA was precipitated at -20°C for 30 min. Precipitations were spun at 14,000 g at 4°C for 15 min, and the supernatant was decanted.

DNA pellets were washed with 80 and 70% ethanol pre-cooled at -20°C, respectively, and spun for an additional 10 min at 4°C. The supernatant was discarded and the DNA pellet was air dried. The pellet was re-suspended in 50 µl nuclease-free H<sub>2</sub>O and kept at 4°C overnight before being stored at -20°C.

## Genome and metagenome sequencing

A total of 99 honeybee gut samples were used for metagenome sequencing (Supplementary Table 1). In total, 83 representative core bacterial strains obtained from *A. cerana* were selected and sequenced to construct a reference genome library for phylotype, SDP, and single nucleotide variation (SNV) analyses (Supplementary Table 2). DNA samples were paired-end sequenced at the Beijing Genomics Institute Shenzhen Branch (BGI-Shenzhen) using the BGISEQ-500 platform (200–400 bp insert size; 100 bp read length; paired-ended [PE]) and at Novogen company using the Illumina Hiseq X Ten platform (350 bp insert size; 150 bp read length; PE). One *Gilliamella* strain (B3022) was sequenced on the PacBio RS II platform at NextOmics company.

## Bacterial genome assembly and annotation

Low-quality reads from the Illumina Hiseq X Ten platform were filtered out using fastp (Chen et al., 2018) (version 0.13.1, -q 20 -u 10) before subsequent analyses. For isolated bacterial strains, clean data were assembled using SOAPdenovo (Luo et al., 2012) (version 2.04, -K 51 -m 91 -R for PE 150 reads; -K 31 -m 63 -R for PE 100 reads), SOAPdenovo-Trans (Xie et al., 2014) (version 1.02, -K 81 -d 5 -t 1 -e 5 for PE 150 reads; -K 61 -d 5 -t 1 -e 5 for PE 100 reads), and SPAdes (Bankevich et al., 2012) (version 3.13.0, -k 33,55,77,85) based on contigs assembled by SOAPdenovo (only for PE 150 reads) or SOAPdenovo-Trans. The assembly with the longest N50 was retained for each strain as the draft genome. Then clean reads were mapped to the assembled scaffolds using minimap 2-2.9 (Li, 2018) and the bam files were generated by samtools (Li H. et al., 2009) (version 1.8). Genome assemblies were processed by BamDeal<sup>1</sup> (version 0.19) to calculate and visualize the sequencing coverage and GC content of the assembled scaffolds. Scaffolds with aberrant depths and GC contents were then removed from the draft genome. Next, the remaining scaffolds were filtered taxonomically. Scaffolds assigned to eukaryote by Kraken2 (Wood et al., 2019) using the standard reference database were removed, and the ones aligned to a wrong phylum by blastn (megablast with  $e < 0.001$ ) were further removed. The remaining genome assemblies were

<sup>1</sup> <https://github.com/BGI-shenzhen/BamDeal>

used as bacterial genome references. The *Gilliamella* strain (B3022) sequenced on the PacBio RS II platform was assembled using a hierarchical genome assembly method (HGAP2.3.0) (Chin et al., 2013).

The protein coding regions of bacterial genomes were predicted using Prokka version 1.13 (Seemann, 2014). The KEGG orthologous groups (KOs) annotation was carried out using KofamKOALA (Aramaki et al., 2020) based on profile HMM and adaptive score threshold with default parameters. Programs KEGG Pathway and Brite Hierarchy were used to screen the annotation results. Finally, dbCAN2 version 2.0.11 (Zhang et al., 2018) was applied to annotate CAZymes and CAZyme gene clusters (CGCs) using embedded tools HMMER, DIAMOND, and Hotpep with default parameters.

## Genetic variation of *A. cerana* hosts

Metagenomes were filtered by fastp (-q 20 -u 10) (Chen et al., 2018). Clean reads were then mapped to the *A. cerana* reference genome (ACSNU-2.0, GCF\_001442555.1) (Park et al., 2015) using the BWA-MEM algorithm (v 0.7.17-r1188) (Li and Durbin, 2010), with default settings and an additional “-M” parameter to reach compatibility with Picard. Read duplicates were marked using Picard MarkDuplicates 2.18.9<sup>2</sup>. GATK HaplotypeCaller in the GVCF mode (McKenna et al., 2010) (v4.0.4) was used to call variants for each sample. All of the per-sample GVCFs were joined using GenotypeGVCFs. Then, the final variant file retained SNPs that met all of the following criteria: (1) average depth >5× and <40×; (2) quality score (QUAL) > 20; (3) average genotype quality (GQ) > 20; (4) minor allele frequency (MAF) > 0.05; (5) proportion of missing genotypes < 50%; and (6) bi-allelic SNP sites.

The identity by state (IBS) distance matrices were performed and constructed with the filtered SNPs using functions “snpgdsIBS” in the R package SNPRelate (Zheng et al., 2012). A neighbor-joining tree was reconstructed based on the IBS distance matrix using the function “nj” in the R package Ape (Paradis et al., 2004). Node support values were obtained after 1,000 bootstrap replicates.

## Reference-based metagenome composition analyses

Shotgun reads generated from the whole honeybee gut were first mapped against the *A. cerana* genome (ACSNU-2.0, GCF\_001442555.1) using BWA aln (version 0.7.16a-r1181, -n 1) (Li and Durbin, 2010) to identify host reads, which were subsequently excluded. For taxonomic assignments of

bacterial sequences, we used Kraken2 (Wood et al., 2019) and Bracken version 2.0 (Lu et al., 2017) to profile bacterial phylotype composition and used MIDAS (Nayfach et al., 2016) to profile strain composition for metagenomic samples. The reference database contained 390 bacterial genomes, including 307 published genomes and 83 newly-sequenced *A. cerana*-derived strains from this study (Supplementary Table 2). The majority of the reference strains belonged to six core phylotypes (*Gilliamella*, *Snodgrassella*, *Bifidobacterium*, *Lactobacillus* Firm-4, *Lactobacillus* Firm-5, and *Apibacter*) of honeybee gut bacteria. The analyses of public gut metagenome data of *A. cerana* from Japan (Ellegaard et al., 2020) and *A. mellifera* (Ellegaard and Engel, 2019; Ellegaard et al., 2020) followed the same pipeline.

## Identification and profiling of sequence-discrete population

We defined SDPs for each core gut bacterium (*Gilliamella*, *Snodgrassella*, *Bifidobacterium*, *Lactobacillus* Firm-4, *Lactobacillus* Firm-5, and *Apibacter*) using a 95% gANI threshold (Richter and Rosselló-Móra, 2009). Pairwise average nucleotide identities were calculated using the pyani Python3 module<sup>3</sup>. To generate the whole-genome tree for each core bacterium, we used Roary version 3.12.0 (Page et al., 2015) with the parameter -blastp 75 to obtain core single-copy genes shared among all strains. The alignments of nucleotide sequences were concatenated, from which a maximum-likelihood tree was inferred using FastTree version 2.1.10 (Price et al., 2010) with a generalized time-reversible (GTR) model and then visualized using iTOL (Letunic and Bork, 2019).

We used the ‘run\_midias.py species’ script of MIDAS (Nayfach et al., 2016) with default parameters to estimate SDP relative abundances in each sample. The script ‘merge\_midias.py species’ with the option ‘-sample\_depth 10.0’ was used to merge SDP abundance files across samples. The SDPs with a relative abundance of less than 1% were filtered out.

## Detection of single nucleotide variation and copy number variations across populations

CheckM version 1.0.86 (Parks et al., 2015) was used to estimate the completeness and contamination of genomes. The genome with the highest completeness and lowest contamination was selected as the reference sequence for each SDP. The metagenomic reads were mapped against reference genomes and the SNVs were quantified along the entire genome using MIDAS (Nayfach et al., 2016) and the

<sup>2</sup> <http://broadinstitute.github.io/picard/>

<sup>3</sup> <https://github.com/widowquinn/pyani>

script 'run\_midass.py snps' with default parameters. For each SDP, the script 'merge\_midass.py snps' pooled data across multiple samples with options '-snp\_type bi -site\_depth 5 -site\_prev 0.05 -sample\_depth 5.0 -fract\_cov 0.4 -allele\_freq 0.01' to obtain the minor allele (second most common) frequency file. Thus, bi-allelic SNVs prevalent in more than 5% of profiled samples were predicted and rare SNVs with abnormally high read depth were excluded. The matrix files of SNVs remaining polymorphic were obtained after filtering steps.

We used the 'run\_midass.py genes' script in MIDAS (Nayfach et al., 2016) to map metagenomic reads to pangenomes of each SDP and quantified gene copy numbers with default parameters. Then, we merged results from pangenome profiling across samples with the option '-sample\_depth 5.0' from the 'merge\_midass.py genes' module. The gene coverage was normalized by the coverage of a set of 15 universal marker genes to obtain the estimated copy number for genes of each SDP. The coverage of each KO term was obtained by summing up all genes annotated as the same KO for each SDP. *p*-values were calculated using the Kruskal–Wallis one-way analysis across populations with the 'compare\_means' function in the R package 'ggpubr.' KO copy number variation and SNV of each SDP were detected as highly variable when an adjusted *p* < 0.05.

## De novo assembly of metagenomes

The metagenome was also *de novo* assembled using MEGAHIT (Li et al., 2016) (version 1.1.2, -m 0.6 -k-list 31,51,71 -no-mercy) for each gut sample. Assemblies longer than 500 bp were blasted against the NCBI nr database using DIAMOND (Buchfink et al., 2015) (version 0.9.22.123, blastx -f 102 -k 1 -e 1e-3) and were assigned to fungi, bacteria, archaea, virus, or plants (Viridiplantae). Only assemblies assigned as bacteria were retained for further analyses.

A customized bacterial genome database was constructed to enable taxonomic assignments for the bacterial assemblies. The database included all bacterial genomes available on NCBI<sup>4</sup> up to Jan 2019 (167,172 genomes), 83 genome assemblies of newly sequenced *A. cerana* gut bacteria (Supplementary Table 2), and 14 *Apibacter* genomes from *A. cerana* (Zhang et al., 2022). Taxonomical assignments were conducted using blastn, and an *e*-value of 1e-5 was observed. The assemblies were assigned to the genus of the best hit, while those without any hits were defined as unassigned bacteria.

For each metagenome sample, all clean reads were mapped against bacterial assemblies using SOAPaligner (Li R. et al., 2009) (version 2.21, -M 4 -l 30 -r 1 -v 6 -m 200). The results were summarized using the soap.coverage script (version

2.7.7<sup>5</sup>). Only assemblies with  $\geq 90\%$  coverage were considered true bacteria. Shannon index and Bray–Curtis dissimilarity were calculated using the vegan R package (Philip, 2003). The analyses of public gut metagenome data of *A. cerana* from Japan (Ellegaard et al., 2020) and *A. mellifera* (Ellegaard and Engel, 2019) followed the same pipeline.

## Gene prediction and functional annotation for metagenomes

Gene prediction was conducted using MetaGeneMark (Zhu et al., 2010) (GeneMark.hmm version 3.38) with the *de novo* metagenome assemblies, and those longer than 100 bp were clustered using CD-HIT (Li and Godzik, 2006) (version 4.7, -c 0.95 -G 0 -g 1 -aS 0.9 -M 0) to obtain a non-redundant gene catalog for *A. cerana* metagenomes. For each individual metagenome sample, clean data were aligned onto the non-redundant gene catalog using SOAPaligner (Li R. et al., 2009) (version 2.21, -M 4 -l 30 -r 1 -v 6 -m 200). The gene abundance was calculated using soap.coverage script (version 2.7.9, see Text Footnote 5). For each sample, only assemblies of  $\geq 90\%$  coverage were retained for further annotation. The analyses of public gut metagenome data of *A. cerana* from Japan (Ellegaard et al., 2020) and *A. mellifera* (Ellegaard and Engel, 2019) followed the same pipeline.

Functional annotation of the gene catalog was performed by GhostKOALA (Kanehisa et al., 2016) using the genus\_prokaryotes KEGG GENES database and KofamKOALA (Aramaki et al., 2020) with an *e*-value threshold of 0.001. Genes were first assigned with KO ID predicted by KofamKOALA, and the remaining unassigned genes were then annotated using GhostKOALA. KOs were mapped onto KEGG pathways using the KEGG Mapper online<sup>6</sup>.

The abundances of KOs and pathways were calculated as the sum of the abundances of all genes annotated to them using custom scripts. Population dissimilarities (Bray–Curtis distance) of KO function among the 15 bee populations were tested by the ANOSIM test included in the vegan package (Philip, 2003) with 999 permutations. Linear discriminant analysis (LDA) was performed using LEfSe (Segata et al., 2011) with default parameters to identify KO biomarkers in different populations. Function enrichment of featured KOs was estimated by one-sided Fisher's Exact Test using the stats R package at both module and pathway levels.

For each featured KO, the abundances for all bacterial species encoding the KO-related genes were listed for all of the 99 samples. In each population, the median abundance was used as the abundance of bacterial species encoding the respective KO. Then, the contributions by different

<sup>4</sup> <https://ftp.ncbi.nlm.nih.gov/genomes/refseq/>

<sup>5</sup> <https://github.com/aquaskyline/SOAPcoverage>

<sup>6</sup> [https://www.kegg.jp/kegg/tool/map\\_pathway2.html](https://www.kegg.jp/kegg/tool/map_pathway2.html)

bacterial species to the corresponding KO were estimated. If the KO term was identified in > 50% of individual bee guts of the same population, the KO was considered to be present in the population. To compare the gene numbers among different populations, we standardized metagenome data by randomly extracting 400 Mb bacterium-derived data from each gut sample, which were mapped to the gene assemblies. The assemblies were retained only if the coverage was  $\geq 90\%$ .

The glycoside hydrolase (GH) and polysaccharide lyase (PL) genes were functionally assigned to the dbCAN2 database (Zhang et al., 2018). In each population, the median abundance was used as the abundance of bacterial species encoding respective GH/PL gene clusters. Then, the contributions by different bacterial species to the corresponding GH/PL gene clusters were estimated.

## Diet profiling of gut and honey metagenomes

A customized chloroplast genome database was first constructed for flowering plants (4,161 from NCBI and 271 newly sequenced ones generated by our group) for KrakenUniq version 0.5.5 (Breitwieser et al., 2018). For gut metagenome data, we filtered out reads mapped to the *A. cerana* genome or to the *de novo* bacterial assemblies and used the remaining reads for pollen diet profiling based on chloroplast DNA found in the gut annotated at the family level. The remaining reads were first aligned to the customized chloroplast genomes with KrakenUniq (Breitwieser et al., 2018) with default parameters. Those mapped reads were aligned to nt database with blastn with an e-value setting as  $1e-5$ , and the best alignment was retained. Then, the reads from the alignments with similarity >95% and query coverage >90% to reference sequences from plants were kept and used to estimate the pollen abundance at the family level. The families with a relative abundance of less than 1% were filtered out.

The geographical variation in pollen composition was also conducted with the assembled metagenome data from honey samples collected from five representative regions of this study (SC\_AB, SC\_GB, SX\_QL, QH\_GD, JL\_DH) (Liu et al., 2021). The assemblies with similarity >95% and query coverage >90% to reference plant sequences were retained. Clean reads were then aligned to these assemblies using Minimap2 (Li, 2018), and the mapped reads were used to estimate the pollen abundance at the family level with SamBamba (Tarasov et al., 2015). The families with a relative abundance of less than 1% were filtered out.

The gut bacterial phylotype and KO composition from *de novo* assembly and annotation were used in the correlation analysis with pollen composition at the family level.

## Heritability of bacterial diversity

The rank-based inverse normal transformation of the relative abundance with the reference-based method was used in the heritability analysis. The heritability was defined as the Percentage of Variance Explained (PVE) and estimated with Genome-wide Efficient Mixed Model Association (GEMMA, v0.94) (Zhou and Stephens, 2012). To control the effects of individual relatedness, population structure, and diet variation, we regressed the transformed gut bacteria abundance with the first three PCs from the PCA of the host genotypic data and the pollen Shannon index from the gut. Then, PVE estimation was performed with the residuals using GEMMA (with relatedness matrices and the HE regression algorithm). A phylotype or SDP was considered heritable if the PVE measurements did not show overlaps with zero.

## The association between host genetic variation and bacterial diversity

The rank-based inverse normal transformation of the relative abundance of core gut bacteria was used in the Genome-Wide Association Studies (GWAS) analysis. We used the Linear Mixed Model in rMVP v1.0.0 (Yin et al., 2021). In the GWAS analysis, the kinship between individuals, the first three PCs in host PCA, and the diet (Shannon index of pollen family composition) were used for correction. We used the 'EMMA' method to analyze variance components in rMVP. The statistical significance level was set to  $p < 5 \times 10^{-8}$  for the GWAS association.

## The effects of diet on the abundance of *Gilliamella* and *Lactobacillus* Firm-5

A *Gilliamella* strain (B2889, belonging to the SDP Gillia\_Acer\_2) was cultivated with HIA, and a *Lactobacillus* Firm-5 strain (B4010) was cultivated with MRS with 0.02 g/ml D-fructose (aladdin F108331) and 0.001 g/ml L-cysteine (aladdin C108238). The microbiota-free *A. cerana* workers were obtained following Zhang et al. (2022). Pupae in the late stage were removed from brood frames and incubated in sterile plastic bins at 35°C. Both bacterial strains of  $OD_{600} = 1$  were mixed at equal volumes and then mixed with 50% (v/v) sterilized sucrose syrup, which was fed to newly emerged microbiota-free honeybees. After 3 days, cellobiose (Shanghai Yuanye Bio-Technology Co., Ltd. S11030, final concentration 5 mg/ml) and solutions with different proportions of pectin (Sigma, P9135) and cellulose (Megazyme, P-CMC4M) (1:10, 10:1, final mixed concentration 5.5 mg/ml) mixed with sterilized 50% sucrose syrup were fed to honeybees, respectively. Honeybees fed with only 50% sucrose syrup were used as control. After feeding for

4 days, DNA was extracted from bee guts and used for the qPCR assay.

## qPCR assay

We conducted real-time qPCR experiments to determine bacterial loads for both *Gilliamella* and *Lactobacillus* Firm-5 after the feeding experiments. 16S-F-Gillia (5'-TGAGTGCTTGCACCTTGATGACG-3') and 16S-R-Gilla (5'-ATATGGGTTTCATCAAATGGCGCA-3') primers were used for *Gilliamella* 16S rRNA gene amplification. 16S-F-Firm5 (5'-GCAACCTGCCCTWTAGCTTG-3') and 16S-R-Firm5 (5'-GCCCCATCCTKTAGTGACAGC-3') primers (Kešnerová et al., 2017) were used for *Lactobacillus* Firm-5 16S rRNA gene amplification. Actin-AC-F (5'-ATGCCAACACTGTCTTCT-3') and Actin-AC-R (5'-GACCCACCAATCCATACGGA-3') primers were used to amplify the *actin* gene of the host *A. cerana* (Park et al., 2020), which was used to normalize the bacterial amplicons (Kešnerová et al., 2017). The 16S target sequences were cloned into vector pEASY-T1 (Transgen), and the Actin target sequence was cloned into pCE2 TA/Blunt-Zero Vector (Vazyme), then confirmed by Sanger sequencing. The copy number of the plasmid was calculated, serially diluted, and used as the standard. qPCR was performed using the ChamQ Universal SYBR qPCR Master Mix (Vazyme) and QuantStudio 1 (Thermo Fisher) in a standard 96-well block (20- $\mu$ l reactions; incubation at 95°C for 3 min, 40 cycles of denaturation at 95°C for 10 s, and annealing/extension at 60°C for 20 s). The data were analyzed using the QuantStudio Design & Analysis Software v1.5.1 (Thermo Fisher) and Excel (Microsoft). *p*-values were calculated using the Mann-Whitney test.

## Results

### Bacterial composition significantly varied across Asian honeybee populations at multiple levels

A total of 99 nurse bees from 15 geographic populations covering 13 provinces across China were analyzed (Figure 1A and Supplementary Table 1). SNPs derived from honeybee reads were used to construct a neighbor-joining tree (Figure 1B), which confirmed the geographic origin of the sampled populations. This result was consistent with the reported genetic structure and geographic distribution of *A. cerana* populations (Ji et al., 2020), thereby excluding the possibility of colony translocation.

Bacterial reads were then *de novo* assembled and aligned against the GenBank nr database to recover the phylotype composition for individual nurse bees. In congruence with

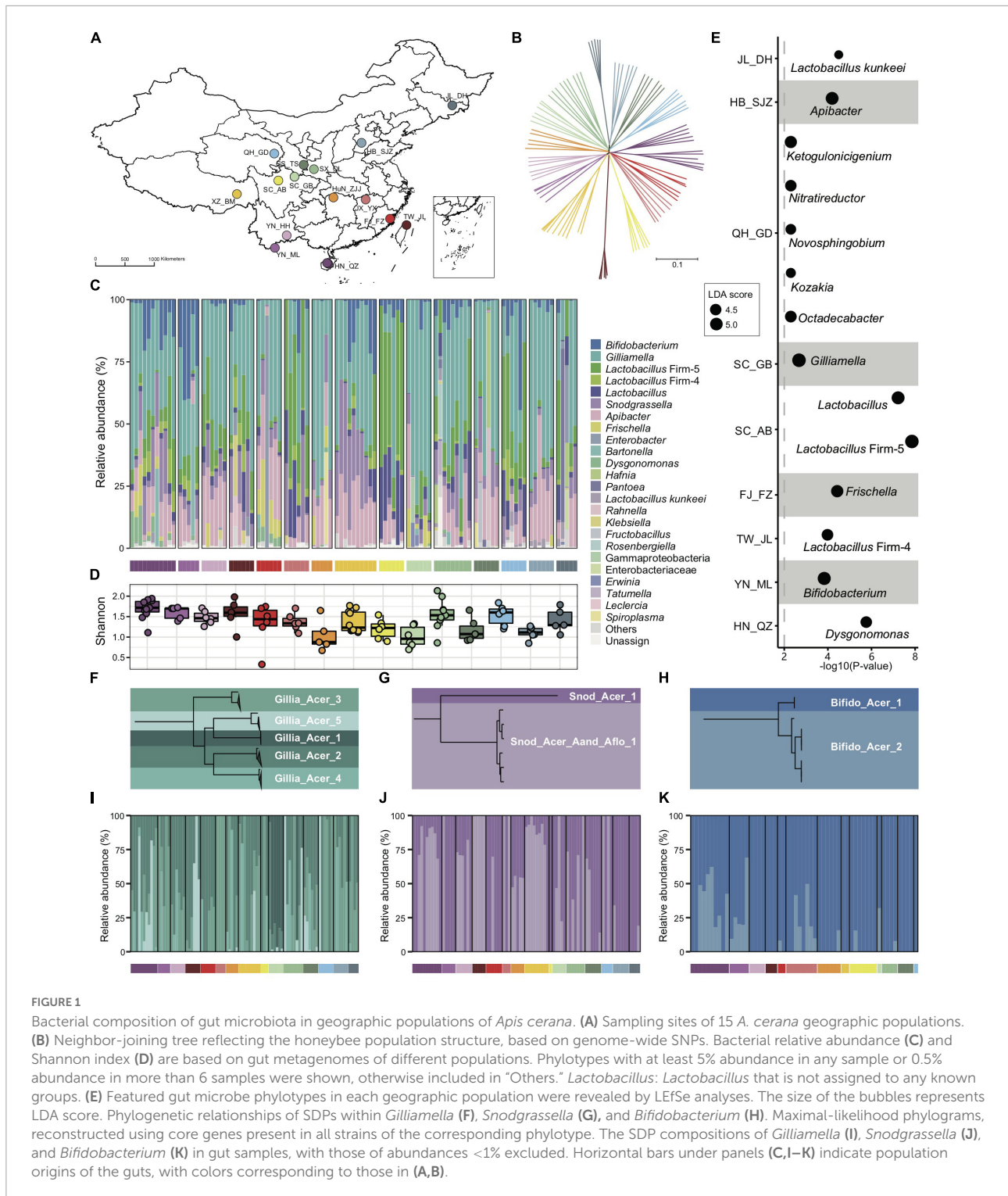
previous studies (Kwong et al., 2017b; Ellegaard et al., 2020), the core gut microbiota in *A. cerana* included six groups of bacteria, i.e., *Gilliamella* and *Snodgrassella* from Proteobacteria, *Bifidobacterium* from Actinobacteria, *Lactobacillus* Firm-4 and Firm-5 from Firmicutes, and *Apibacter* from Bacteroidetes (Figure 1C). This result was further confirmed by the reference-based method (Supplementary Figures 1, 2), which employed a customized database containing 307 public and 83 newly sequenced bee gut bacterial genomes (Supplementary Table 2). However, apparent variations in phylotype composition were observed among individual bees (Figure 1C), and the composition of core microbes appeared to be less stable than in *A. mellifera* (Regan et al., 2018; Ellegaard and Engel, 2019; Ge et al., 2021).

Both Shannon index (Figure 1D, Kruskal-Wallis,  $P = 0.0022$ ) and phylotype diversity (ANOSIM,  $r = 0.29$ ,  $P = 0.001$ ) showed noticeable differences across populations of *A. cerana*. Nine of the 15 investigated populations could be defined by featured bacteria in the LEfSe analysis (Segata et al., 2011), which showed significantly higher relative abundances in the focal population than all remaining populations (Figure 1E).

The distinct gut variation across host populations was also reflected at finer taxonomic scales. Among all six core phylotypes in *A. cerana*, *Gilliamella* contained the most diverse host-specific sequence-discrete populations (SDPs) (Figures 1F-H and Supplementary Figures 3-6), which were defined as strains sharing a genome-wide average nucleotide identity (gANI) >95% within each phylotype. Our results revealed varied presence and abundance in SDPs of core phylotypes among gut samples (Figures 1I-K and Supplementary Figure 7), whereas *Gilliamella* showed significant SDP differences among geographical populations (Supplementary Figure 8, ANOSIM  $r = 0.14$ ,  $p = 0.001$ ). We also identified genome sites showing single nucleotide variation (SNV) for major SDPs in each sample, to detect gut variations at the strain level (Supplementary Figure 9). The results demonstrated significant variations in SNV composition across populations (Supplementary Figure 10). Thus, the gut bacterial composition of Asian honeybees varied significantly across geography at phylotype, SDP, and strain levels.

### Progressive changes in the honeybee microbial community were related to diet shift

Gut compositions showed extensive overlaps among populations, forming continuous groups in PCoA analyses (Figure 2A), indicating progressive changes in microbial community structure among endemic honeybee populations. Interestingly, a continuous variable contributing to the separation along the first principal coordinate axis (PCoA) reflected antagonistic dynamics in abundances of *Gilliamella*



and *Lactobacillus* Firm-5 (Figure 2B). Among all six core phylotypes, the relative abundance of *Gilliamella* (Spearman's  $\rho = -0.85, p = 2.14e-28$ ) and *Lactobacillus* Firm-5 (Spearman's  $\rho = 0.79, p = 4.47e-22$ ) showed the most significant correlation with the PCoA1 value.

In the 99 samples, the populations from XZ\_BM, SC\_AB, HN\_QZ, JL\_DH, TW\_JL, and QH\_GD represent each of the peripheral six subspecies, and the others are from the central ancestral population (Ji et al., 2020). We first tested whether the gut compositions showed differences at the subspecies level. We



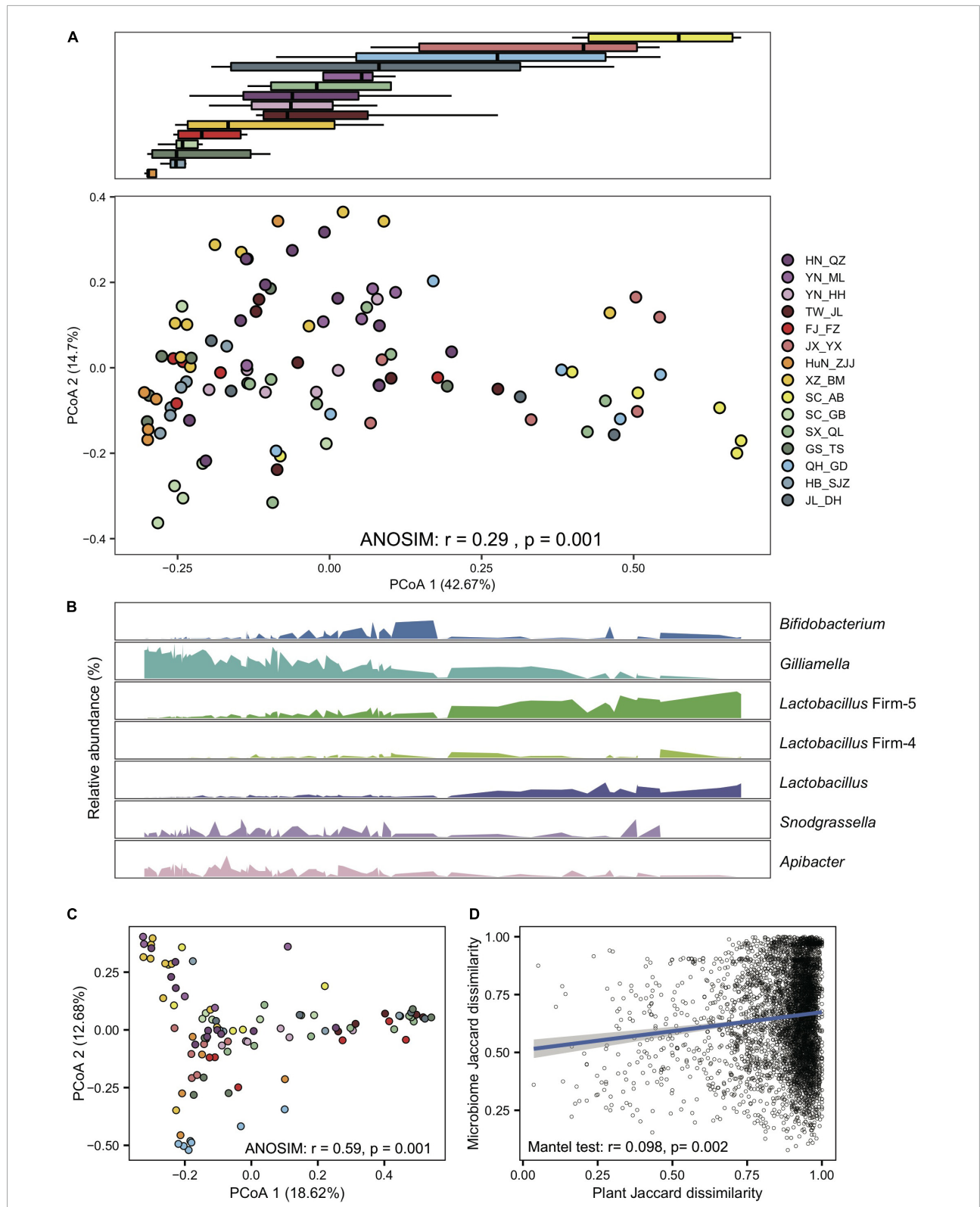


FIGURE 2

*Gilliamella* and *Lactobacillus* Firm-5 showed antagonistic trends in compositional turnover of honeybee gut microbiomes. (A) Overall variation of the gut microbial community at the phylotype level, revealed by Bray–Curtis dissimilarity PCoA (bottom panel). Boxplots (top panel) indicate the distribution of each population along the first principal coordinate axis (PCoA1). Boxplot center values represent the median and error bars represent the SD. Colors correspond to the population origin of the gut samples. (B) Relative abundances of core bacterial phylotypes along PCoA1. (C) The pollen composition at the family level varied in gut metagenomes from populations of *A. cerana*. (D) The Jaccard distances of the gut bacterial phylotype and the pollen composition at the family level were significantly correlated.

compared the Bray–Curtis distance between each of the central populations and between central and peripheral populations. The gut variations among subspecies were significantly more prominent than those within subspecies at the phylotype level (Wilcoxon test,  $p = 1.4e-7$ ) but not at the SDP level (Wilcoxon test,  $p = 0.87$ ). The results indicated that the gut microbiome changed along host differentiation, which might be related to the host genetic differentiation and diet variation.

Next, we estimated the heritability of the relative abundance of core bacteria at both phylotype and SDP levels. The heritability was overall low. Among the core phylotypes, *Gilliamella* abundance showed the highest heritability (Supplementary Figure 11), while that of *Snodgrassella* was not obvious. The abundances of about one-third SDPs were not heritable. The GWAS analysis did not detect any apparent site variation that had determined bacterial composition, as no genomic region of *A. cerana* was found significantly associated with the bacterial abundance (with a threshold as  $p < 2e-8$ ) at either the phylotype or SDP level. These results indicated that gut microbial diversity at the population level was not likely driven by single-site nucleotide variations. The complex genetic heterogeneity and limited sample size might also mask the effect of host genetics.

To examine the effect of diet on the gut microbiome, we first extracted pollen reads from the metagenome data and identified flower composition for each gut sample (details in Section “Materials and methods”). Honeybee populations from different regions showed significant variation in pollen diet at the family level (ANOSIM,  $r = 0.59$ ,  $P = 0.001$ , Figure 2C and Supplementary Table 3). Such a diet shift was further confirmed by pollen variation in honey samples extracted from five of the representative populations (SC\_AB, SC\_GB, SX\_QL, QH\_GD, and JL\_DH), where pollen composition again showed significant differences at the family level (ANOSIM,  $r = 0.35$ ,  $p = 0.007$ , Supplementary Figure 12 and Supplementary Table 4). Most importantly, the Jaccard distances of the gut bacterial phylotype and the pollen composition were significantly correlated (mantel test,  $r = 0.098$ ,  $p = 0.002$ , Figure 2D). Among the core phylotypes, the abundances of *Gilliamella* showed a significant correlation with the Shannon index of pollen composition in the gut (Spearman’s  $\rho = -0.23$ ,  $p = 0.020$ ). Therefore, the pollen diet showed a correlation with the composition of the honeybee gut microbiome.

## KEGG orthology function was correlated with diets and characterized by carbohydrate metabolism and transport

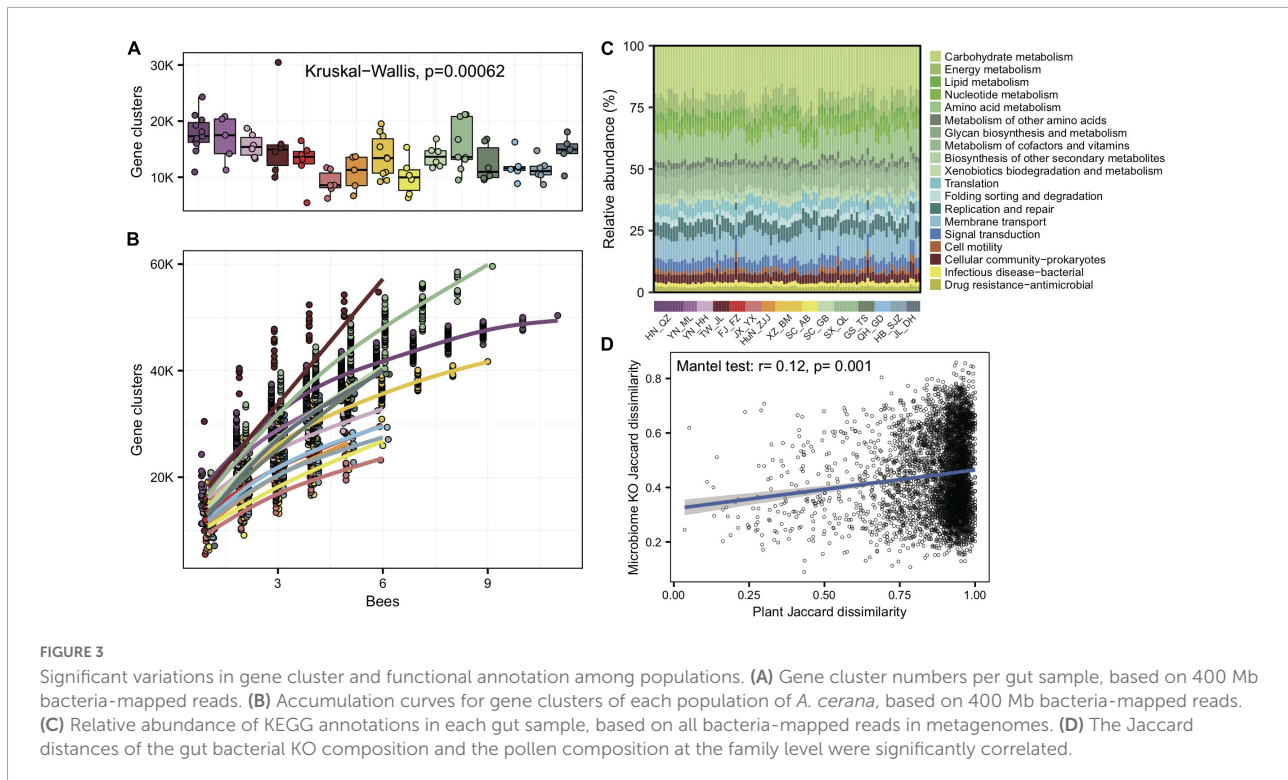
To understand whether gut microbes in *A. cerana* showed idiosyncratic regional traits on the function level, we *de novo*

assembled the metagenomes and annotated genes for each of the 99 gut samples. As with bacterial compositions, the number of gene clusters per gut varied significantly among populations (Kruskal–Wallis test,  $p = 6.2e-4$ ) (Figure 3A). The gene cluster number in different individuals was significantly correlated with the Shannon index of gut bacteria (Pearson’s  $r = 0.64$ ,  $p = 8.28e-13$ ), suggesting that bacterial diversity is the basis for gene varieties among individual bees. We also quantified the rate of novel gene accumulation for each population. The results demonstrated distinct differences in the genetic diversity among populations (Figure 3B).

We assigned predicted gene clusters from all metagenome data to the KEGG database to reveal the diversity of functions among populations. A total of 1,965 functional orthologs (KOs) shared among all populations were enriched in genetic information processing and signaling and cellular processes (Supplementary Figure 13). The KO category compositions (Figure 3C) also showed extensive overlap and were distinctively differentiated among populations (ANOSIM,  $r = 0.33$ ,  $p = 0.001$ , Supplementary Figure 14). The LEfSe analyses showed that 11 of the 15 geographic populations had noticeably enriched KO categories (Supplementary Figure 15), which showed significantly higher relative abundances in the focal population than all remaining populations.

We also tested whether the KO compositions showed a difference at the subspecies level. The gut bacterial KO composition among subspecies was significantly more prominent than those within subspecies (Wilcoxon test,  $p = 1.6e-4$ ). Furthermore, the Jaccard distances of the gut bacterial KO composition and pollen diversity at the family level showed a significant correlation (mantel test,  $r = 0.12$ ,  $p = 0.001$ , Figure 3D). The results indicated that not only bacterial composition but also their functions changed along host differentiation and were associated with diets.

The top population-enriched KOs ( $p < 1e-5$ ) mainly included functions in metabolism and membrane transport (Supplementary Figure 15). At the KO term level, we identified 83 KO terms showing inter-population differences (Supplementary Table 5), in which they were significantly more abundant in only one of the geographic populations. Interestingly, 37 of 83 of the enriched KO terms were transporter pathway genes (all belonging to ko02000) (Figure 4A), whereas the pathway was also enriched in some local populations (e.g., SC\_AB and YN\_ML, Figure 4B). Most featured transporters were related to carbohydrates (Figure 4C) and six of the enriched KO terms belonged to the glycoside hydrolase (GH) family (Supplementary Table 5), in concert with the fact that polysaccharides are one of the major nutritional components derived from pollen. Therefore, the population-enriched gut microbe KOs were mainly associated with carbohydrate metabolism and transport and were significantly correlated to pollen composition in a given local environment.



## Phosphotransferase system, ATP binding cassette transporters, and glycoside hydrolases contributed by *Gilliamella*, *Lactobacillus* Firm-5 and *Bifidobacterium* were hotspot genes involved in local adaptation

In congruence with the finding that carbohydrate metabolism and transport play important roles in adapting to local diets, key genes of such pathways, such as phosphotransferase system (PTS) transporters and ATP binding cassette (ABC), were often characterized in distinct honeybee populations. For instance, a total of 17 PTS and 16 ABC transporters were identified from the 37 enriched transporter pathway genes (Supplementary Table 5). All featured PTS genes were only found in the SC\_AB population, while the featured ABC transporters were present in several populations (XZ\_BM, SC\_AB, and YN\_ML). PTS serves as one of the major mechanisms in carbohydrate uptake, particularly for hexoses and disaccharides. In SC\_AB, the 17 featured PTS genes included some that are specific for ascorbate, beta-glucoside, cellobiose, fructoselysine/glucoselysine, galactitol, mannose, and sucrose (Supplementary Table 5). The mapping of relevant gene clusters against the bacterial nr database suggested that these featured PTS genes were mainly contributed by *Gilliamella* and *Lactobacillus* Firm-5 (Supplementary Table 6). The dominant role of these two bacteria in coding PTS genes

was further confirmed by analyses of 81 individually sequenced and annotated genomes, where *Gilliamella* and *Lactobacillus* Firm-5 were the major phylotypes encoding PTS genes (Supplementary Table 7). At the SDP level, *Lactobacillus* Firm-5 had a higher copy number of PTS transporters for cellobiose, fructoselysine/glucoselysine, and galactitol than *Gilliamella* (Figure 5A). Many of these PTS transporters were found in the featured genes in the SC\_AB population, which was dominated by *Lactobacillus*. Thus, the enrichment of featured PTS genes could at least be partially explained by the elevated abundance of the contributing bacteria in local populations (Figure 1E).

The featured ABC transporters included transporters for amino acids, iron, and carbohydrates (Supplementary Table 5). Besides *Gilliamella* and *Lactobacillus* Firm-5, *Bifidobacterium* also contributed unique ABC transporters (Supplementary Table 6). For example, the *Bifidobacterium*-unique transporters for raffinose/stachyose/melibiose (msmE, msmF, and msmG) (genome annotation results in Supplementary Table 7) were featured in the YN\_ML population, in which *Bifidobacterium* was also the featured phylotype (Figure 1E). The elevated *Bifidobacterium* and its unique ABC transporters characterized in YN\_ML might be attributed to the presence of raffinose and stachyose in specific pollen or nectar, which are toxic to the honeybees (Barker, 1977).

At a finer taxonomic scale, 14 of the 17 featured PTS genes had significant population-distinct SNV sites coded by SDP from *Lactobacillus* Firm-5, and 9 of the 16 ABC transporters

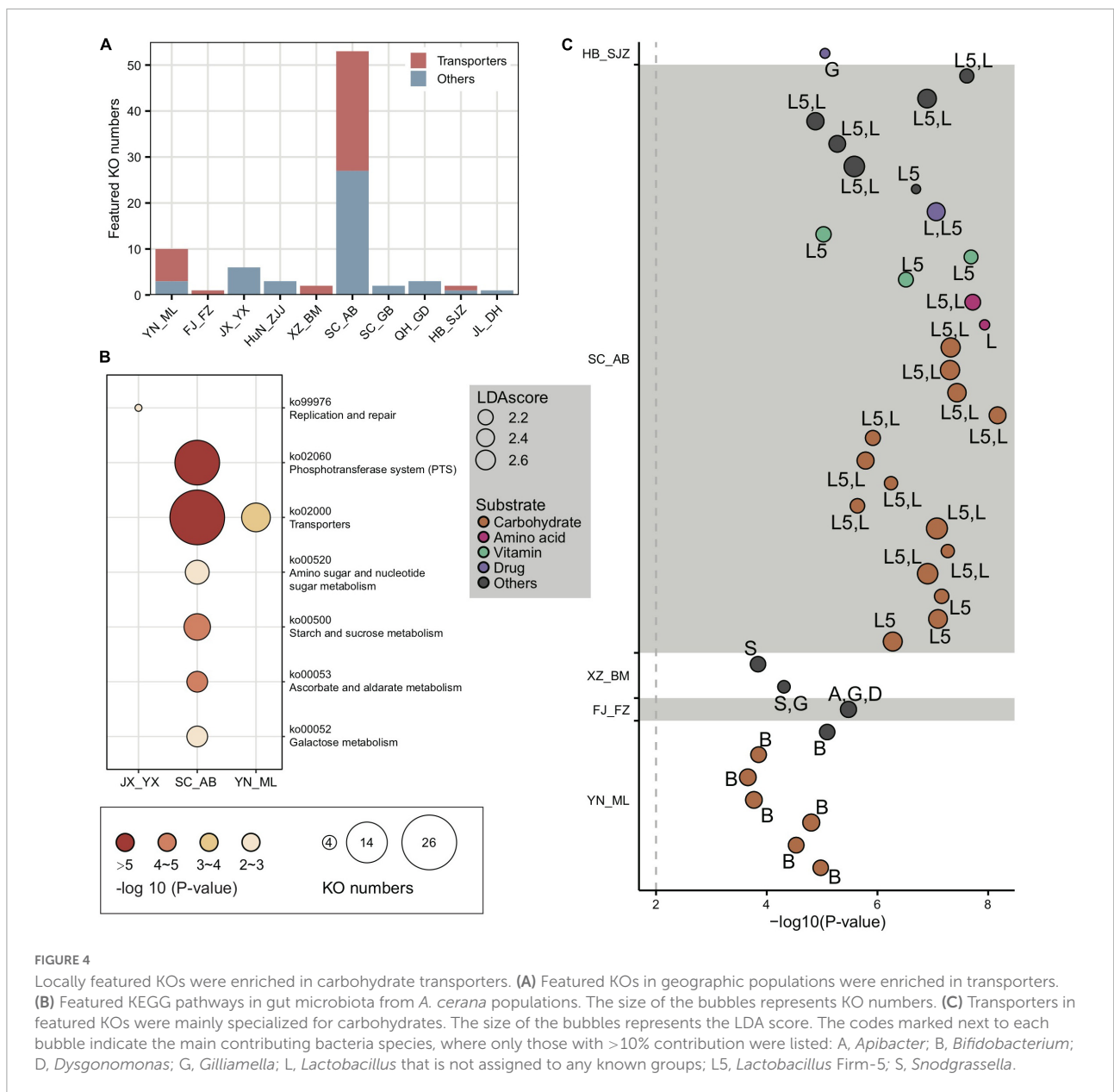


FIGURE 4

Locally featured KOs were enriched in carbohydrate transporters. (A) Featured KOs in geographic populations were enriched in transporters. (B) Featured KEGG pathways in gut microbiota from *A. cerana* populations. The size of the bubbles represents KO numbers. (C) Transporters in featured KOs were mainly specialized for carbohydrates. The size of the bubbles represents the LDA score. The codes marked next to each bubble indicate the main contributing bacteria species, where only those with >10% contribution were listed: A, *Apibacter*; B, *Bifidobacterium*; D, *Dysgonomonas*; G, *Gilliamella*; L, *Lactobacillus* Firm-5; S, *Snodgrassella*.

harbored significant population-distinct SNV sites coded by SDPs from *Lactobacillus* Firm-5 and *Apibacter* (Supplementary Table 8). One featured gene *ulaC* (ascorbate PTS system EIIA or EIIAB component, K02821), coded by SDP from *Lactobacillus* Firm-5 showed significant population-distinct copy number variations (CNVs) (Supplementary Table 9). Thus, the variations in functional genes seemed to have been caused by changes in the featured bacterial composition at both phylotype and strain levels.

Besides PTS and ABC genes, six GH genes were featured in *A. cerana* populations (from GH1, GH3, GH29, GH36, GH43, and GH78 family) and were mainly contributed by *Gilliamella*, *Lactobacillus*, and *Bifidobacterium* (Figure 5B

and Supplementary Table 10). To construct the profile for major gene families involved in glycoside breakdown, we used dbCAN2 (Zhang et al., 2018) to annotate all GH and polysaccharide lyase (PL) genes. We discovered that the GH/PL profiles varied across populations (Supplementary Figure 16). Additionally, the non-core bacterium also encoded novel 1265 GH genes. For instance, *Dysgonomonas* contributed unique GH gene families in *A. cerana*, including GH57, GH92, GH133, and GH144 (Supplementary Table 10). This non-core bacterium was featured in the HN\_QZ population (Figure 1E), likely due to its contribution to unique GH gene sets.

Some of the six featured GH genes were positioned together with featured PTS or ABC transporters on the genome.

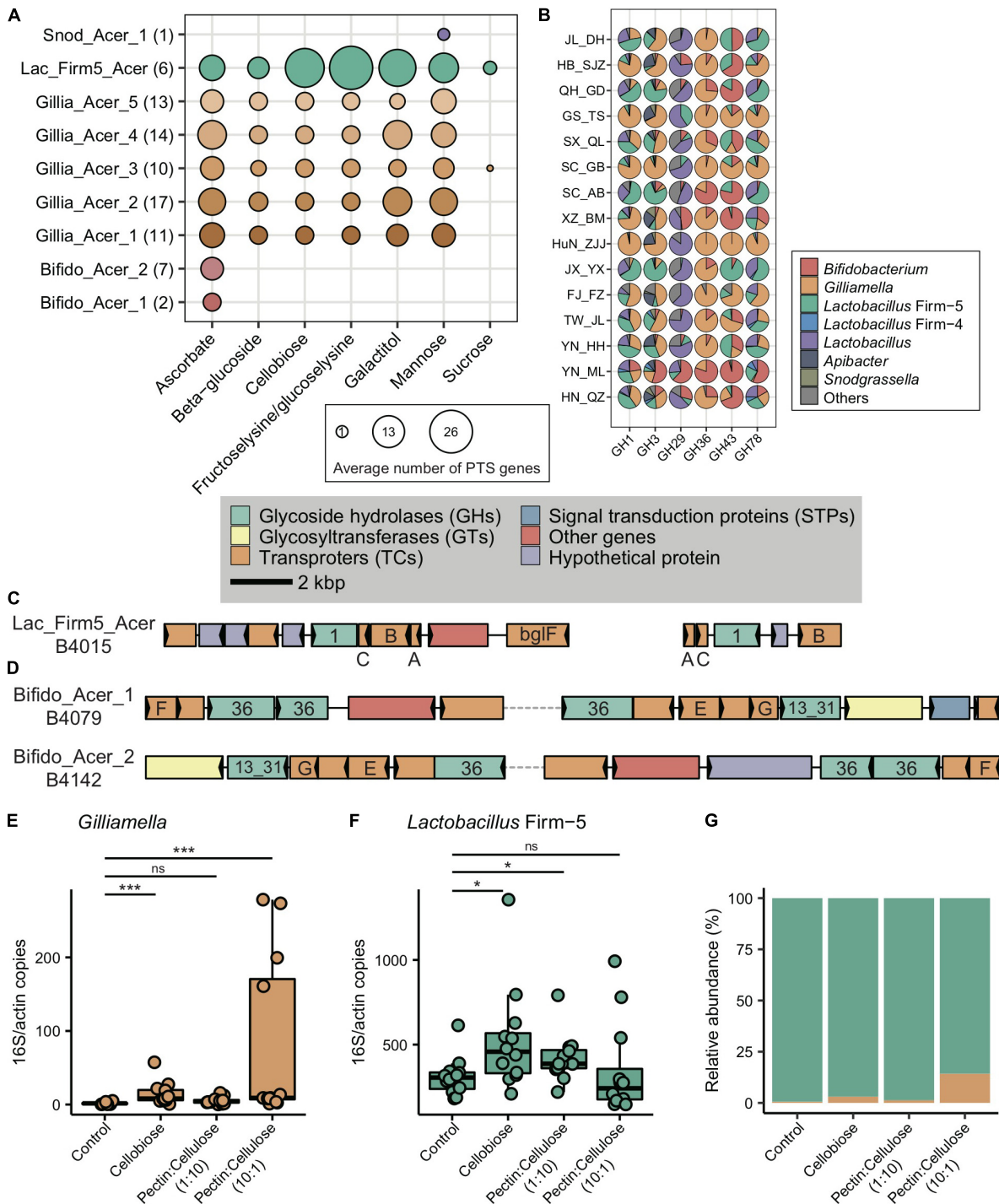


FIGURE 5

Main bacterial phylotypes coding for PTS and GHs. (A) Gene copy numbers in population-featured PTS pathways identified in all SDPs. Numbers in parentheses represent SDP strain numbers. Genes were annotated from the genomes of newly isolated microbial strains from *A. cerana* guts. (B) Featured GHs were coded by different bacterial phylotypes from metagenome of 15 geographic populations of *A. cerana*. (C) PTS transporters (*celA/celB/celC/bglF*), 6-phospho-beta-glucosidase (*bglA*) from the GH1 family were often found located together in genomes, which were represented here by *Lactobacillus Firm-5* SDP. (D) ABC transporters (*msmE/msmF/msmG*), alpha-galactosidase from the GH36 family, and alpha-glucosidase from the GH13\_31 family were often found located together in genomes, which were represented here by two *Bifidobacterium* SDPs. The change in the absolute abundance of *Gilliamella* (E), *Lactobacillus Firm-5* (F), and the percentage of *Gilliamella* and *Lactobacillus Firm-5* (G) after feeding cellobiose and mixtures of pectin and cellulose with different concentrations. PTS, phosphotransferase system; GH, glycoside hydrolase. ns, not significantly different, \* $p < 0.05$ , \*\*\* $p < 0.001$ .

Together, these genes formed CAZyme gene clusters (CGCs), performing sequential functions in polysaccharide degradation and transportation. For example, in *Lactobacillus* Firm-5, the featured 6-phospho-beta-glucosidase (*bglA*) from the GH1 family, PTS system genes for beta-glucoside, and cellobiose were usually clustered and formed CGCs (Figure 5C), and all these genes were enriched in the SC\_AB population. In *Bifidobacterium*, the raffinose/stachyose/melibiose transport system *msmEFG* and alpha-galactosidase from the GH36 family involved in raffinose/melibiose degradation were usually located together (Figure 5D). These genes were all featured in the YN\_ML population, which had *Bifidobacterium* as the featured phylotype.

## The feeding experiment verified the contribution of pollen polysaccharide composition to the trade-off of *Gilliamella* and *Lactobacillus* Firm-5

Our investigation of *A. cerana* guts from its natural range revealed antagonistic abundance between the two core-bacteria *Gilliamella* and *Lactobacillus* Firm-5 across geographic populations (Figure 2B). As both lineage and function diversities of honeybee gut bacteria were correlated to pollen diets (Figures 2D, 3D), we speculate that characteristic traits in local food resources may have led to bacterial community shifts observed at the grand scale. To test this hypothesis, we conducted feeding experiments to verify whether functional adaptations observed in metagenomes can lead to adaptive advantages in bacterial competition.

As the main structural components of the pollen wall, pectin and cellulose were chosen as representative nutritional contents to examine the impacts of food on the abundance variation between *Gilliamella* and *Lactobacillus* Firm-5 in co-feeding experiments. In the main gut microbe phylotypes in the honeybee, only *Gilliamella* are able to degrade the polygalacturonic acid (PGA), the backbone of pectin (Engel et al., 2012). On the other hand, cellobiose (the key metabolite of cellulose) related PTS genes (Supplementary Table 5) and the metabolic pathway (*ko00500*, starch, and sucrose metabolism) were highly enriched in the SC\_AB population, as revealed by the metagenome data. The newly assembled *Lactobacillus* Firm-5 genome also showed elevated copy numbers in cellobiose PTS (Figure 5A). As such, we anticipated that local food with a higher proportion of pectin would increase the fitness of *Gilliamella*, and food with a higher proportion of cellulose would favor *Lactobacillus* Firm-5 in the community.

We fed *A. cerana* workers that were colonized with an equal abundance of *Gilliamella* and *Lactobacillus* Firm-5, with cellobiose, pectin, and cellulose mixture with different concentrations (1:10 and 10:1, respectively) and examined corresponding changes in bacterial composition after 4 days.

Interestingly, the absolute abundance of *Lactobacillus* Firm-5 was always higher than *Gilliamella* in the control group, which was only fed with sucrose (Figures 5E–G), indicating a predominant role of *Lactobacillus* over *Gilliamella* in the given condition. The absence of pollen in food, and the absence of sucrose PTS genes in the strain we used (belonging to *Gillia\_Acer\_2* SDP, Figure 5A) might explain the low abundance of *Gilliamella* in the control group. The absence of *Snodgrassella* might also affect the colonization of *Gilliamella* (Martinson et al., 2012; Kwong et al., 2014).

After feeding cellobiose, the absolute abundance of both *Gilliamella* and *Lactobacillus* Firm-5 significantly increased relative to the control group (Figures 5E,F), which was in accordance with the presence of cellobiose PTS genes in both phylotypes (Figure 5A). As expected, *Gilliamella* and *Lactobacillus* Firm-5 showed different responses to the mixed food with varied concentrations of pectin and cellulose. The absolute abundance of *Gilliamella* did not show significant variation after feeding food of pectin:cellulose (1:10), but the abundance of *Lactobacillus* Firm-5 significantly increased (Figures 5E,F). On the other hand, the absolute abundance of *Gilliamella* showed a significant increase after feeding food of pectin:cellulose (10:1), but the abundance of *Lactobacillus* Firm-5 did not vary significantly (Figures 5E,F). The varied proportion of pectin and cellulose impacted the antagonistic of *Gilliamella* and *Lactobacillus* Firm-5. These results suggested that diet, pollen polysaccharide, in particular, is an important factor in shaping gut bacterial composition and functions in *A. cerana*.

## Discussion

### The progressive change of gut microbiome in Asian honeybee populations

In this study, we carried out comprehensive investigations on the gut microbiomes of the widespread Asian honeybee *A. cerana* at the population level. While many studies have contributed to our knowledge of the honeybee gut microbiota, little is understood about how this essential symbiont system evolves with the host. In agreement with previous studies on both *A. mellifera* (Rothman et al., 2018; Ellegaard and Engel, 2019) and *A. cerana* (Ge et al., 2021), our results revealed variations in gut microbes among *A. cerana* individuals, even among those from the same hive. The intra-colony variation might be related to differed social interactions among honeybee individuals (Powell et al., 2014) or varied exposure to the stored pollens and other hive materials (Anderson et al., 2022) in the honeybee.

Our studied system involved 15 geographic populations of *A. cerana*, and we were particularly interested in understanding

gut variations among the seven genetically distinctive populations, which we showed in our recent study (Ji et al., 2020) that had diverged genetically and morphologically at a subspecies level. These subspecies have been confined to drastically different habitats (e.g., mountain valleys of >3,000 m, tropical islands, temperate plains, hills, etc.). In contrast to the abrupt distinction between *A. mellifera* and *A. cerana*, the gut microbiome of honeybee populations showed progressive change within *A. cerana* (Supplementary Figure 17). The bacterial compositions across populations showed significant variations at phylotype, SDP, strain, and gene content levels, albeit with extensive overlaps. The strain composition of *Gilliamella* and *Snodgrassella* was largely similar among populations of western honeybees from four different states in the United States (Bobay et al., 2020). The gut microbiota community from 18 different human populations across geography also showed extensive overlap (Smits et al., 2017), implying a common trend in gut microbiome evolution for hosts exhibiting a continuous and wide-range distribution.

In the western honeybee, host genetics influenced the gut microbe composition, where *Bifidobacterium* abundance was associated with the genotype of the host glutamate receptor (Wu et al., 2021). Different from Wu et al. (2021), the complex background heterogeneity combined with a limited sample size might have masked apparent host genetic influence on gut bacteria at the local scale in our study, which may explain the weak signal reported in our GWAS analysis. However, our findings genuinely reflect the host genetic background and associated microbiota, which could not have been discovered without a broad scale sampling. With an in-depth understanding of the molecular mechanisms underlying honeybee host-gut bacteria interactions in the future, we expect that a more focused genomic screening on these target genes would reveal their specific contributions in widely distributed native bee populations.

## Gut microbiome evolution under local diet shift in Asian honeybee

The honeybees consume relatively simple but consistent food, i.e., pollen and honey, and pollen is especially important to the gut microbes. Controlled experiments on the diet with or without pollen influenced the total and specific gut bacteria abundance (Kešnerová et al., 2020; Ricigliano and Anderson, 2020). Pollen diet also facilitates the co-existence of closely related *Lactobacillus* species by using different pollen-derived carbohydrate substrates (Brochet et al., 2021). Although controlled experiments conducted on *A. mellifera* have built the foundation on diet influence on honeybee gut microbiota, we knew little about how natural diets influenced honeybee gut microbiota in their native range.

Floral shifts are a common theme during range expansion and habitat adaptation of the honeybees (Ji et al., 2020). The change of locally flowering plants inevitably alters nutrients for honeybees and the associated gut microbiome, because nutritional components vary in both pollen and nectar across plant species. Pollen walls are enriched in polysaccharides in the forms of cellulose, hemicelluloses, and pectin, which serve as major food resources for the gut bacteria (Engel et al., 2012; Zheng et al., 2019). Previous studies have shown the contents of cellulose and hemicellulose varied in pollen of different species (McLellan, 1977) and in bee pollen collected from different regions (Herbert and Shimanuki, 1978). Similarly, sugar composition in nectar varied among flowers (Chalcoff et al., 2006). Particularly, nectar may contain low doses of sugars that are toxic to the honeybee, such as raffinose and mannose (Barker, 1977). Thus, both the general floral configuration and specific flower traits could serve as determining factors for the formation of a local honeybee gut profile.

Our recent work on the evolution of mainland *A. cerana* revealed that the changing floras led to a convergent adaptation of the honeybee (Ji et al., 2020). Here, we showed that both microbial composition and function of the honeybee gut microbiota exhibited progressive change throughout the studied natural range. The variation could be partially explained by the pollen diet, which is closely related to changing flora in the habitat. Such an intra-species transition in the gut microbiome reflects the evolutionary consequence of collective adaptation of both the honeybee and its symbionts.

Besides amino acids, lipids, and vitamins, pollen is a source of diverse carbohydrate sources. Carbohydrate metabolism is the second most abundant functional class of bacterial transcripts (Lee et al., 2015). Different honeybee gut bacteria species showed varied GH transcripts (Lee et al., 2015) and activities (Ricigliano et al., 2017). The PTS and ABC transporters, genes involved in the transportation of multiple types of polysaccharides, were also associated with different gut bacteria species (Lee et al., 2015). Accordingly, the PTS and ABC transporters were primarily encoded by *Gilliamella* and *Lactobacillus* Firm-5, representing the most enriched transporters among all bacterial genes featured in local populations of *A. cerana*. Our feeding and inoculation assays further showed that pollen polysaccharides determined the abundance of the two core bacteria, *Gilliamella* and *Lactobacillus* Firm-5. The role of core bacteria in local adaptations was reinforced by evidence showing their dominant contributions to genes related to pollen and nectar digestions.

Unexpectedly, non-core bacteria sometimes became abundant in local honeybee populations. For instance, *Dysgonomonas* was typically low in abundance among *A. cerana* individuals, as reported in both *Apis nigrocincta* (Lombogia et al., 2020) and *A. mellifera* (Erban et al., 2017). But this bacterium contributed a series of unique GH genes in FJ\_FZ and HN\_QZ populations, thereby becoming abundant

and common in the corresponding gut microbiome. This observation suggested that local food resources might trigger bacterial species turnover when non-core bacteria became more suited to new diets, which again highlights the significance of diet on the gut profile.

## Population heterogeneity needs to be considered for the evolution and adaptation of honeybee microbiomes

A recent study suggested that both lineage and function diversities of the gut microbes were significantly lower in *A. cerana* when compared with *A. mellifera* (Ellegaard et al., 2020). However, this conclusion was drawn based on two *A. mellifera* colonies from Switzerland, two colonies of both *A. mellifera* and *A. cerana* from two sites in Japan, it is difficult to anticipate whether such a distinct pattern could be generalized when population gradients of both honeybee species are taken into consideration. Although the present study was not designed for comprehensive analyses of inter-species comparisons, our results provided insights into how intra-species variations in gut microbiota might affect interpretations of differences between honeybee species.

Although the per-bee gene diversity was generally lower in *A. cerana* microbiota than *A. mellifera*, individual bees of different *A. cerana* populations showed variation (Supplementary Figure 18A). In addition, the divergence of the accumulated gene diversity between the two species was much less significant than previously suggested. The Japanese populations representing *A. cerana* in the earlier study (Ellegaard et al., 2020) were one of the least variable populations among all *A. cerana* populations investigated in this study (Supplementary Figure 18B). Given the large variations observed among *A. cerana* populations, it is unknown whether a similar difference might also be common within *A. mellifera* and how that might influence the distinctions between these two widely distributed honeybee species. Additionally, other confounding factors should also be taken into consideration to gain a comprehensive understanding of honeybee gut microbiomes. In particular, the evolutionary pathways and phylogenetic relationships of focal populations, the specifics in honeybee management (such as colony merging and artificial diet additions), and other human interventions, may all have significant impacts on the honeybee gut profile. As the gut symbiont profile is a signature of the natural adaptation of the host to specific habitats, it would seem that comparisons between microbiomes of intra- and inter-host honeybee species should always be placed in a context of specific environments.

Since worker age (Hroncova et al., 2015) and seasonality (Almeida et al., 2022) showed effects on honeybee gut microbiome, these factors need to be considered in the data interpretation. Additionally, the limited sampling for each local

population might also under-estimate the gut microbe diversity and bring bias into intra-colony variation (Rothman et al., 2018). However, season control and simultaneous age marking are admittedly difficult for sampling honeybees from a wide geographic range. In our study, we chose a practical method to specifically sample nurse bees based on their morphology. Admittedly, such criteria are not as accurate as individual marking and errors are possible. Despite the potential influence of seasonality, populations sampled in the same month did not show any elevated affinity to those sampled from different months (Figure 2 and Supplementary Figure 14). Nevertheless, with honeybee age and season controlled, collecting enough individual bees at the intra- and inter- colony levels could improve future research initiatives on investigations of native honeybee colonies. Notably, season control is to collect bees with the same circadian activity and probably not the same month for different populations located across temperate and tropical regions.

Our study took a first step toward understanding the relative contribution of diet and host genetics on the gut microbiota of widely ranged honeybee populations. Our results detected localized gut features at both species and functional levels throughout the distribution range. However, the gut microbiome showed unexpected extensive overlap across the investigated ranges, which covered temperate to tropical regions. These results suggested that progressive change is the foundation of gut microbiome evolution in the Asian honeybee and specialized bacterial traits help to adapt to local diets. In this regard, regional floral diversity could serve as a key to maintaining characteristic repertoires of honeybee gut microbes, which is tremendously important for honeybee health as a whole. Therefore, a sustaining plant community containing diverse endemic flower species should be considered a key part of a honeybee conservation plan. On the other hand, the fitness of gut microbiomes of the honeybee populations may play an unforeseen role in the survival of colonies, during honeybee introduction, hybridization, and especially translocation.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ncbi.nlm.nih.gov/>, PRJNA705951.

## Author contributions

XiZ designed the study. XiZ, SL, and XeZ organized and coordinated the study. QS coordinated sample collection, bacterial isolation, and genome annotation, reference-based metagenome mapping, and SDP analysis. MT conducted



*de novo* assembly of bacteria and metagenome and metagenomics analysis. SL conducted genetic variation analysis of *A. cerana*, heritability, and GWAS analysis. JH conducted diet profiling. QS and JT conducted feeding experiments and qPCR assay. XiZ, QN, and XL organized sampling. SL, XiZ, and XgZ wrote the first draft. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Beijing Natural Science Foundation (No. 5204035) and the National Natural Science Foundation of China (No. 32000343) to SL, the Program of the Ministry of Science and Technology of China (2018FY100403), National Special Support Program for High-level Talents (Ten-Thousand Talents Program), the National Natural Science Foundation of China (No. 31772493), and funding from the Beijing Advanced Innovation Center for Food Nutrition and Human Health through China Agricultural University grant to XiZ, the National Natural Science Foundation of China (No. 31470123) to XL, and the 2115 Talent Development Program of China Agricultural University through XiZ.

## Acknowledgments

Many colleagues and collaborators have facilitated sampling or contributed samples to this work, including Kang Lai, Yongkun Ji, Jieluan Li, Dandan Lang, Shuang Yang, Shijie Liang, Zhanhai Cai, Dingqian Xiang, Jie Deng, Hua Liang, Kangkai Zheng, Abuduoji, Liang Li, Jihuan Du, Zijing Zhang, Xuhui Wang, Zhaohuai Li, Feixiang Wang, Yanqiong Peng, Jialong

## References

- Almeida, E. L., Ribiere, C., Frei, W., Kenny, D., Coffey, M. F., and O'toole, P. W. (2022). Geographical and seasonal analysis of the honeybee microbiome. *Microbial. Ecol.* [Epub ahead print]. doi: 10.1007/s00248-022-01986-x
- Anderson, K. E., Ricigliano, V. A., Copeland, D. C., Mott, B. M., and Maes, P. (2022). Social interaction is unnecessary for hindgut microbiome transmission in honey bees: The effect of diet and social exposure on tissue-specific microbiome assembly. *Microbial. Ecol.* [Epub ahead print]. doi: 10.1007/s00248-022-02025-5
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859
- Baldo, L., Riera, J. L., Tooming-Klunderud, A., Albà, M. M., and Salzburger, W. (2015). Gut microbiota dynamics during dietary shift in Eastern African cichlid fishes. *PLoS One* 10:e0127462. doi: 10.1371/journal.pone.0127462
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021
- Barker, R. J. (1977). Some carbohydrates found in pollen and pollen substitutes are toxic to honey bees. *J. Nutr.* 107, 1859–1862. doi: 10.1093/jn/107.10.1859
- Bobay, L. M., Wissel, E. F., Raymann, K., and Campbell, B. J. (2020). Strain structure and dynamics revealed by targeted deep sequencing of the honey bee gut microbiome. *mSphere* 5:e694–e620. doi: 10.1128/mSphere.00694-20
- Breitwieser, F. P., Baker, D. N., and Salzberg, S. L. (2018). KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* 19:198. doi: 10.1186/s13059-018-1568-0
- Brochet, S., Quinn, A., Mars, R. A. T., Neuschwander, N., Sauer, U., and Engel, P. (2021). Niche partitioning facilitates coexistence of closely related honey bee gut bacteria. *eLife* 10:e68583. doi: 10.7554/eLife.68583
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Chalcoff, V. R., Aizen, M. A., and Galletto, L. (2006). Nectar concentration and composition of 26 species from the temperate forest of South America. *Ann. Bot.* 97, 413–421. doi: 10.1093/aob/mcj043
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. doi: 10.1093/bioinformatics/bty560

Huang, and Zhengwei Wang. Chenyi Li, Yating Du, Chengfeng Yang, Lizhen Guo, Lifei Qiu, and Wenjun Zhang from China Agricultural University isolated and cultured honeybee gut bacteria that were used for genome sequencing in this work. We thank Chengfeng Yang from China Agricultural University for providing the primers and plasmid of *Gilliamella* used for qPCR assay. We also thank Hao Zheng from China Agricultural University for his comments on the early drafts of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.934459/full#supplementary-material>

- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474
- Cryan, J. F., and Dinan, T. G. (2012). Mind-altering microorganisms: The impact of the gut microbiota on brain and behaviour. *Nat. Rev. Neurosci.* 13, 701–712. doi: 10.1038/nrn3346
- Ellegaard, K. M., and Engel, P. (2019). Genomic diversity landscape of the honey bee gut microbiota. *Nat. Commun.* 10:446. doi: 10.1038/s41467-019-08303-0
- Ellegaard, K. M., Suenami, S., Miyazaki, R., and Engel, P. (2020). Vast differences in strain-level diversity in the gut microbiota of two closely related honey bee species. *Curr. Biol.* 30, 2520–2531. doi: 10.1016/j.cub.2020.04.070
- Engel, P., James, R. R., Koga, R., Kwong, W. K., McFrederick, Q. S., and Moran, N. A. (2013). Standard methods for research on *Apis mellifera* gut symbionts. *J. Apicultural Res.* 52, 1–24. doi: 10.3896/IBRA.1.52.4.07
- Engel, P., Martinson, V. G., and Moran, N. A. (2012). Functional diversity within the simple gut microbiota of the honey bee. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11002–11007. doi: 10.1073/pnas.1202970109
- Engel, P., and Moran, N. A. (2013). The gut microbiota of insects – diversity in structure and function. *FEMS Microbiol. Rev.* 37, 699–735. doi: 10.1111/1574-6976.12025
- Erban, T., Ledvinka, O., Kamler, M., Hortova, B., Nesvorna, M., Tyl, J., et al. (2017). Bacterial community associated with worker honeybees (*Apis mellifera*) affected by European foulbrood. *PeerJ.* 5:e3816. doi: 10.7717/peerj.3816
- Ge, Y., Jing, Z., Diao, Q., He, J. Z., and Liu, Y. J. (2021). Host species and geography differentiate honeybee gut bacterial communities by changing the relative contribution of community assembly processes. *mBio* 12:e0075121. doi: 10.1128/mBio.00751-21
- Henderson, G., Cox, F., Ganesh, S., Jonker, A., Young, W., and Janssen, P. H. (2015). Rumen microbial community composition varies with diet and host, but a core microbiome is found across a wide geographical range. *Sci. Rep.* 5:14567. doi: 10.1038/srep14567
- Herbert, E. W., and Shimanuki, H. (1978). Chemical composition and nutritive value of bee-collected and bee-stored pollen. *Apidologie* 9, 33–40. doi: 10.1051/apido:19780103
- Hroncova, Z., Havlik, J., Killer, J., Duskocil, I., Tyl, J., Kamler, M., et al. (2015). Variation in honey bee gut microbial diversity affected by ontogenetic stage, age and geographic location. *PLoS One* 10:e0118707. doi: 10.1371/journal.pone.0118707
- Ji, Y., Li, X., Ji, T., Tang, J., Qiu, L., Hu, J., et al. (2020). Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Sci. Adv.* 6:eabd3590. doi: 10.1126/sciadv.abd3590
- Kamada, N., Chen, G. Y., Inohara, N., and Núñez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. *Nat. Immunol.* 14, 685–690. doi: 10.1038/ni.2608
- Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Kešnerová, L., Emery, O., Troilo, M., Liberti, J., Erkosar, B., and Engel, P. (2020). Gut microbiota structure differs between honeybees in winter and summer. *ISME J.* 14, 801–814. doi: 10.1038/s41396-019-0568-8
- Kešnerová, L., Mars, R. A. T., Ellegaard, K. M., Troilo, M., Sauer, U., and Engel, P. (2017). Disentangling metabolic functions of bacteria in the honey bee gut. *PLoS Biol.* 15:e2003467. doi: 10.1371/journal.pbio.2003467
- Kwong, W. K., Engel, P., Koch, H., and Moran, N. A. (2014). Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11509–11514. doi: 10.1073/pnas.1405838111
- Kwong, W. K., Mancenido, A. L., and Moran, N. A. (2017a). Immune system stimulation by the native gut microbiota of honey bees. *R. Soc. Open Sci.* 4:170003. doi: 10.1098/rsos.170003
- Kwong, W. K., Medina, L. A., Koch, H., Sing, K. W., Soh, E. J. Y., Ascher, J. S., et al. (2017b). Dynamic microbiome evolution in social bees. *Sci. Adv.* 3:e1600513. doi: 10.1126/sciadv.1600513
- Kwong, W. K., and Moran, N. A. (2016). Gut microbial communities of social bees. *Nat. Rev. Microbiol.* 14, 374–384. doi: 10.1038/nrmicro.2016.43
- Lee, F. J., Rusch, D. B., Stewart, F. J., Mattila, H. R., and Newton, I. L. (2015). Saccharide breakdown and fermentation by the honey bee gut microbiome. *Environ. Microbiol.* 17, 796–815. doi: 10.1111/1462-2920.12526
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47:W256–W259. doi: 10.1093/nar/gkz239
- Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H. F., Sadakane, K., et al. (2016). MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, R., Yu, C., Li, Y., Lam, T. W., Yiu, S. M., Kristiansen, K., et al. (2009). SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966–1967. doi: 10.1093/bioinformatics/btp336
- Li, W., and Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi: 10.1093/bioinformatics/btl158
- Liu, S., Lang, D., Meng, G., Hu, J., Tang, M., and Zhou, X. (2021). Tracing the origin of honey products based on metagenomics and machine learning. *Food Chem.* 371:131066. doi: 10.1016/j.foodchem.2021.131066
- Lombogia, C. A., Tulung, M., Posangi, J., and Tallei, T. E. (2020). Bacterial composition, community structure, and diversity in *Apis nigrocincta* gut. *Int. J. Microbiol.* 2020:8. doi: 10.1155/2020/6906921
- Lu, J., Breitwieser, F. P., Thielen, P., and Salzberg, S. L. (2017). Bracken: Estimating species abundance in metagenomics data. *PeerJ. Comput. Sci.* 3:e104. doi: 10.7717/peerj-cs.104
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., et al. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1:18. doi: 10.1186/2047-217X-1-18
- Martinson, V. G., Moy, J., and Moran, N. A. (2012). Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl. Environ. Microbiol.* 78, 2830–2840. doi: 10.1128/AEM.07810-11
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- McLellan, A. R. (1977). Minerals, carbohydrates and amino acids of pollens from some woody and herbaceous plants. *Ann. Bot.* 41, 1225–1232. doi: 10.1093/oxfordjournals.aob.a085413
- Michel, A. J., Ward, L. M., Goffredi, S. K., Dawson, K. S., Baldassarre, D. T., Brenner, A., et al. (2018). The gut of the finch: Uniqueness of the gut microbiome of the Galápagos vampire finch. *Microbiome* 6:167. doi: 10.1186/s40168-018-0555-8
- Nayfach, S., Rodriguez-Mueller, B., Garud, N., and Pollard, K. S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Res.* 26, 1612–1625. doi: 10.1101/gr.201863.115
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421
- Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412
- Park, D., Jung, J. W., Choi, B.-S., Jayakodi, M., Lee, J., Lim, J., et al. (2015). Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genomics* 16:1. doi: 10.1186/1471-2164-16-1
- Park, M. G., Kim, W. J., Choi, J. Y., Kim, J. H., Park, D. H., Kim, J. Y., et al. (2020). Development of a *Bacillus thuringiensis* based dsRNA production platform to control sacbrood virus in *Apis cerana*. *Pest Manag. Sci.* 76, 1699–1704. doi: 10.1002/ps.5692
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Philip, D. (2003). Computer program review VEGAN, a package of R functions for community ecology. *J. Veget. Sci.* 14, 927–930. doi: 10.1111/j.1654-1103.2003.tb02228.x
- Powell, J. E., Martinson, V. G., Urban-Mead, K., and Moran, N. A. (2014). Routes of acquisition of the gut microbiota of the honey bee *Apis mellifera*. *Appl. Environ. Microbiol.* 80, 7378–7387. doi: 10.1128/AEM.01861-14

- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- Radloff, S. E., Hepburn, C., Randall Hepburn, H., Fuchs, S., Hadisoelilo, S., Tan, K., et al. (2010). Population structure and classification of *Apis cerana*. *Apidologie* 41, 589–601. doi: 10.1051/apido/2010008
- Regan, T., Barnett, M. W., Laetsch, D. R., Bush, S. J., Wragg, D., Budge, G. E., et al. (2018). Characterisation of the British honey bee metagenome. *Nat. Commun.* 9:4995. doi: 10.1038/s41467-018-07426-0
- Richter, M., and Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Ricigliano, V. A., and Anderson, K. E. (2020). Probing the honey bee diet-microbiota-host axis using pollen restriction and organic acid feeding. *Insects* 11:291. doi: 10.3390/insects11050291
- Ricigliano, V. A., Fitz, W., Copeland, D. C., Mott, B. M., Maes, P., Floyd, A. S., et al. (2017). The impact of pollen consumption on honey bee (*Apis mellifera*) digestive physiology and carbohydrate metabolism. *Arch. Insect Biochem. Physiol.* 96:e21406. doi: 10.1002/arch.21406
- Rothman, J. A., Carroll, M. J., Meikle, W. G., Anderson, K. E., and McFrederick, Q. S. (2018). Longitudinal effects of supplemental forage on the honey bee (*Apis mellifera*) microbiota and inter- and intra-colony variability. *Microbial Ecol.* 76, 814–824. doi: 10.1007/s00248-018-1151-y
- Seeley, T. D. (2018). Adaptive significance of the age polyethism schedule in honeybee colonies. *Behav. Ecol. Sociobiol.* 11, 287–293. doi: 10.1007/BF00299306
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12:R60. doi: 10.1186/gb-2011-12-6-r60
- Smits, S. A., Leach, J., Sonnenburg, E. D., Gonzalez, C. G., Lichtman, J. S., Reid, G., et al. (2017). Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* 357, 802–806. doi: 10.1126/science.aan4834
- Sommer, F., and Bäckhed, F. (2013). The gut microbiota — masters of host development and physiology. *Nat. Rev. Microbiol.* 11, 227–238. doi: 10.1038/nrmicro2974
- Suzuki, T. A., and Ley, R. E. (2020). The role of the microbiota in human genetic adaptation. *Science* 370:eaz6827. doi: 10.1126/science.aaz6827
- Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., and Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* 31, 2032–2034. doi: 10.1093/bioinformatics/btv098
- Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., et al. (2014). A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genetics* 46, 1081–1088. doi: 10.1038/ng.3077
- Wood, D. E., Lu, J., and Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome Biol.* 20:257. doi: 10.1186/s13059-019-1891-0
- Wu, J., Lang, H., Mu, X., Zhang, Z., Su, Q., Hu, X., et al. (2021). Honey bee genetics shape the strain-level structure of gut microbiota in social transmission. *Microbiome* 9:225. doi: 10.1186/s40168-021-01174-y
- Wu, Y., Zheng, Y., Chen, Y., Chen, G., Zheng, H., and Hu, F. (2020). *Apis cerana* gut microbiota contribute to host health through stimulating host immune system and strengthening host resistance to *Nosema ceranae*. *R. Soc. Open Sci.* 7:192100. doi: 10.1098/rsos.192100
- Xie, Y., Wu, G., Tang, J., Luo, R., Patterson, J., Liu, S., et al. (2014). SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30, 1660–1666. doi: 10.1093/bioinformatics/btu077
- Yatsunenko, T., Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., et al. (2012). Human gut microbiome viewed across age and geography. *Nature* 486, 222–227. doi: 10.1038/nature11053
- Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., et al. (2021). rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. *Genomics Proteomics Bioinformatics* 19, 619–628. doi: 10.1016/j.gpb.2020.10.007
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46:W95–W101. doi: 10.1093/nar/gky418
- Zhang, W., Zhang, X., Su, Q., Tang, M., Zheng, H., and Zhou, X. (2022). Genomic features underlying the evolutionary transitions of *Apibacter* to honey bee gut symbionts. *Insect Sci.* 29, 259–275. doi: 10.1111/1744-7917.12912
- Zheng, H., Perreau, J., Powell, J. E., Han, B., Zhang, Z., Kwong, W. K., et al. (2019). Division of labor in honey bee gut microbiota for plant polysaccharide digestion. *Proc. Natl. Acad. Sci. U.S.A.* 116, 25909–25916. doi: 10.1073/pnas.1916224116
- Zheng, H., Powell, J. E., Steele, M. I., Dietrich, C., and Moran, N. A. (2017). Honeybee gut microbiota promotes host weight gain via bacterial metabolism and hormonal signaling. *Proc. Natl. Acad. Sci. U.S.A.* 114, 4775–4780. doi: 10.1073/pnas.1701819114
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* 28, 3326–3328. doi: 10.1093/bioinformatics/bts606
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genetics* 44, 821–824. doi: 10.1038/ng.2310
- Zhu, W., Lomsadze, A., and Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* 38:e132. doi: 10.1093/nar/gkq275